Dreaming the Whole Cat: Generative Models, Predictive Processing, and the Enactivist Conception of Perceptual Experience

Citation for published version:
Clark, A 2012, 'Dreaming the Whole Cat: Generative Models, Predictive Processing, and the Enactivist Conception of Perceptual Experience', *Mind*, vol. 121, no. 483, pp. 753-771.
https://doi.org/10.1093/mind/fzs106

Digital Object Identifier (DOI):
10.1093/mind/fzs106

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Mind

Dreaming the Whole Cat: Generative Models, Predictive Processing, and the Enactivist Conception of Perceptual Experience.

ANDY CLARK
University of Edinburgh
andy.clark@ed.ac.uk

Abstract

Does the material basis of conscious experience extend beyond the boundaries of the brain and central nervous system? In Clark 2009 I reviewed a number of 'enactivist' arguments for such a view and found none of them compelling. Ward (2012) rejects my analysis on the grounds that the enactivist deploys an essentially world-involving concept of experience that transforms the argumentative landscape in a way that makes the enactivist conclusion inescapable. I present an alternative (prediction-and-generative–model based) account that neatly accommodates all the positive evidence that Ward cites on behalf of this enactivist conception, and that (I argue) makes richer and more satisfying contact with the full sweep of human experience.

1. Dialogues Concerning Experience and Action

In Clark 2009 I explore and reject a number of arguments meant - or so it seemed - to favour a 'hypothesis of the extended conscious mind' (ECM). ECM, as I characterized it, was the claim that 'the local [material] vehicles of some of our conscious experiences might include more than the whirrings and grindings of the brain/CNS' (Clark 2009 p.967). Despite strongly believing (see Clark 1997) that the best way to understand mind and cognition is by attending to complex webs of structure spanning brain, body, and world, and despite arguing (Clark and Chalmers 1998, Clark 2008, and elsewhere) that certain non-conscious cognitive states might be realized by material circuits that loop through brain, body, and aspects of surrounding 'designer environments', I thus rejected – on empirical grounds - the suggestion that the material vehicles of conscious experience loop actually do outside the brain and central nervous system of the experiencing animal.

In an important and illuminating recent treatment, Ward (2012) claims that my negative assessment of enactivist arguments for ECM is fundamentally misguided, as it fails to recognize the true shape of the enactivist's project. That project, Ward suggests, is not (or not directly) to secure a wider material base for conscious experience so much as to urge upon us an alternative conception of the nature of conscious experience itself: one from which the apparently radical conclusions of ECM flow naturally, perhaps even inexorably. In this way 'once we understand the enactivists as urging a specific personal-level conception of experience we can see why ECM follows from their views' (Ward 2012, p. ??).

Thus restructured (and here compressing much of Ward 2012 section 1) the key moves in the enactivist argument look like this:

> 1. There is no a priori reason to restrict the scope of the material (or, as Ward prefers, the sub-personal) basis of conscious experience to the realm of the neural, as realized by the brain/central nervous system.
>
> 2. Knowledge concerning the correct personal-level conception of experience is relevant to a characterization of its sub-personal basis.
>
> 3. The correct personal-level conception of experience depicts experiences as 'essentially interactions between a subject and parts of the world' (Ward 2012, p. ??)

I happily embrace the first two moves in this argument. Supposing that the third was granted then, Ward argues, ECM is as good as established. For the internalist alternative to ECM (I call my favoured alternative 'internalist' largely for want of a better term: but see sections 5 and 6 following) offers a picture of the sub-personal basis of experience that appears to comport badly with the personal-level conception enshrined in (3) above. As Ward puts it:

> "If we can give an exhaustive subpersonal characterization of my experiential state solely in terms of how things are with me internally then how can that state, when viewed at the personal level, be essentially a dynamic relationship between me and my environment?" (Ward 2012, p. ??)

The upshot, Ward rightly notes, is that 'enactivists think that ECM follows fairly trivially from their position' (Ward 2012, p. ??).

Much turns, then, on that third move: the move to an essentially interactive and world-involving conception of personal-level experience itself. I do think that there is something right about that conception. What is right is the idea that experience, as it unfolds in most normal daily (awake) circumstances, is directly world-revealing, and involves a crucial and complex dance between sensory transduction and real-world action. It is this dance (sometimes referred to as 'animate' or 'active' perception – see e.g. Ballard 1991, Churchland, P.S. et al 1994) that, in such cases, determines the contents of the unfolding experience. But it does not follow that the material (sub-personal processing) basis of that experience must then expand to encompass its own objects. Instead, the material apparatus can still quite reasonably be thought to be wholly internal, consisting in the way neural systems both elicit and respond to signature peturbations from the environment.

I do not (at this point) claim to offer an argument for such a conception. I turn to that matter in sections 3 and 4 below. The immediate point is just that, prima facie, a personal-level conception of experience as genuinely world-revealing and as typically involving a variety of crucial interactions with the environment does not demand the rejection of more internalist sub-personal accounts in anything like the immediate way that Ward suggests[1]. I think Ward is right, however, to suggest that we try considering the enactivist arguments as first and foremost arguments for an alternative conception of experience itself, with the conclusions concerning ECM flowing (or at any rate, being supposed to flow) fairly directly from that conception. What, then, are the arguments meant to suggest this crucial conception?

2. The Missing Arguments: Presence, Plasticity, and Fit

As far as I can see, there are two arguments or bodies of evidence whose role (according to Ward) is best understood as an attempt directly to establish the enactivist conception of experience. These are accompanied by a further consideration concerning the proper fit between the correct personal-level conception and the sub-personal facts. Taken together, these arguments and evidence, accompanied by the notion of proper fit, are to deliver the enactivist conception of *experience itself*, from which a version of ECM is meant trivially to follow. Although I considered aspects of these arguments in Clark 2009, I did not consider them, as Ward rightly notes, as first and foremost arguments for the enactivist conception of experience itself. Let's see how things look once we make this perspectival shift.

[1] In the end, I think that the only reading of the third move that will serve Ward's argumentative purpose will turn out to be one in which that step is read as involving a quite independently motivated claim of metaphysical necessity. Such a reading may be suggested by Ward's apparent endorsement of arguments found in Putnam 1999 and McDowell 1986. I do not believe, however, that the enactivists' argument for ECM was meant simply as a rehearsal of these more familiar philosophical worries thought to afflict an internalist metaphysic. Nor do I believe (see section 5 following) that the particular forms of experiential internalism I defend actually fall foul of such worries. My concern, in any case, was with a suite of apparently quite different arguments and considerations: ones that attempt to provide additional, more empirically-based, reasons for embracing ECM.

The first argument/body-of-evidence involves a complex of claims and ideas concerning 'virtual representation', 'virtual presence', and what Noë sometimes refers to (see e.g. the discussion in Noë 2006 pp. 413-428) as 'presence-in-absence'. The key observation here is that experience presents a world of objects, rather than (merely) a world of (what might be called) 'perspectival takes' on objects. Thus:

> The visual experience of the tomato, when one takes it at face value, presents itself to one precisely as a *visual experience as of a whole tomato*. (Noë 2006, p. 413)

This is not to deny that we also see (for example) one side of the tomato rather than the other. Instead, both kinds of information are somehow present in the visual experience. But this can seem puzzling:

> How can it be true, as I think it is, that we are perceptually aware, when we look at a tomato, of parts of the tomato which, strictly speaking, we do not perceive. This is the puzzle of perceptual presence. (Noë 2006, p. 414)

This is a puzzle about experience. How can experience encompass that which is not, in some sense, immediately transduced? The answer, according to Noë (and endorsed by Ward) is that "Phenomenologically, the world is given in perception as *available*" (Noë 2006, p.422). We have a sense of the visual presence of the whole tomato because we know how to move our eyes and heads to retrieve more information as required, and we know that if we do so, the 'absent' side of the tomato will become viewable, and so on. The same story accounts, Noë claims, for our experience of the full current visual scene as detailed, despite the well-known limits of acuity that characterize peripheral vision. Presence in absence is thus to be explained by presence as access:

> The detail is present *now*, though absent (unseen, out of view, partially occluded etc) because we *now* possess the skills needed to bring the relevant features into view. (Noë 2006, p.422)

This delivers, Ward wants to suggest, the enactivist conception of experience itself. Once we take the story about access as the correct account of the feeling of presence in absence, experience is revealed as a way of (or a potentiality of) acting in the world:

> Qualities are available in experience as possibilities, as potentialities, but not as complete givens. Experience is a dynamic process of navigating the pathways of these possibilities. (Noë 2006 p. 428)

Ward is right to point to these kinds of observation as shedding light on the notion of virtual presence (virtual representation) that I found especially puzzling in Clark 2009. Such observations are indeed meant to convince us that "experience is a mode of temporally extended skilful interaction with the world" (Ward 2012 p. ??). But this is convincing only to the extent that the best explanation of these various facts (concerning presence-in-absence) is indeed the one that conceives of experience itself in this way. In section 3 I shall cast substantial doubt on this implicit suggestion.

The second argument/body-of-evidence meant to speak in favour of the enactivist conception of experience concerns neural plasticity and the so-called 'variable neural correlates argument'. Here too, Ward worries that my previous engagement with the argument (Clark 2009 pp. 969-972) mistakenly treats it as a failed attempt at a direct argument for ECM, rather than a (potentially more successful) argument meant to support the enactivist conception of experience. The key move here consisted in displaying a variety of cases in which the inner story varies greatly, but the agent's experience is plausibly said to remain to some appreciable extent the same. Cases include the successful re-wiring of visual input to auditory cortex in young ferrets (Sur et al 1999, discussed in Noë 2004 p.227) and the (at least partial) successes of sensory substitution technologies. These allow tactile or auditory channels to code inputs tranduced from head-mounted cameras in ways that come, with training and practice, to support visuomotor engagements with the world and can even result in reports of quasi-visual experience (for a fairly recent review, see Bach y Rita and Kercel 2003).

In each case (there are other cases too, but these are representative) the empirical considerations, or so Ward argues, are not meant as direct arguments for ECM. Instead they are best seen, together, as building a case for the enactivist conception of experience itself. Thus if we ask why it is that different transduction modes (the use of tactile or auditory channels to route visual information to the brain) or different neural areas (the use of 'auditory' cortex for vision) can support the same kinds of experience, the answer might be: because the various configurations all come to support the same patterns of interaction with the world, and it is these patterns (not any resultant neural goings-on) that actually constitute the visualness of the experience. As Ward puts it:

> Such cases are intended to provide evidence that sameness of interactive relationship with the world goes hand in hand with sameness of experience. This is supposed to support a conclusion not about the sub-personal but about what experience, for the enactivist, essentially is – a skill-mediated relation to the world. (Ward, 2012, p. ??)

If we follow[2] Ward and the enactivist this far, the considerations of appropriate 'fit' between the personal and the sub-personal adumbrated in section 1 above bite. The sub-personal machinery of experience, if experience *really is* skill-mediated interactive engagement with the actual world, cannot be located in the brain/CNS any more than (to follow Ward's examples from section 4 of his treatment) the machinery of automotion, if automotion *really is*

2

the fluid engagement of car and actual road, could be located in the engine.

However, just as in the previous case (of presence-in-absence) all this is convincing only to the extent that the best explanation of these various facts (in this case, the ones concerning variable neural correlates) is indeed the one that conceives of experience itself in this way. Time, then, to consider an alternative.

3. Generative Models in Perception, Prediction, and Action.

The last decade of research in machine learning and in computational cognitive neuroscience displays a pleasing convergence on an explanatory theme. That theme is the use of generative models, trained and deployed within a broad framework of prediction-driven learning and response, as a general model of cortical sensory processing. There are many important variants, nuances, and outright differences marking the various approaches that share this broadly convergent framework.[3] But for present purposes I shall simply present the barest possible sketch of some of the main common features, before exploring their implications for the enactivist's 'inference to the best explanation' concerning presence-in-absence and variable neural correlates. The key idea is that the brain uses prediction-driven processing routines to acquire and deploy hierarchical generative models of the hidden causes (sometimes called latent variables) that best explain the changing patterns of sensory input that impinge upon the agent. It is worth briefly unpacking each of these notions in turn.

[3] Some key contributions (but note that these cover a range of approaches) include Rao and Ballard 1999; Hinton et al 2006; Hinton 2007a; Friston 2005; Friston et al 2011. For some useful reviews of different aspects of this admittedly complex landscape see Hinton 2007b; Bubic et al 2010; Huang and Rao 2011. Important precursors include Dayan et al 1995; Hinton et al 1994; Hinton and Zemel 1994; Neisser 1967; Gregory 1980. For some more philosophical engagements with aspects of this emerging story, see Grush 1997, 1999, 2004; Clark and Grush 1999; Clark 2006, forthcoming-a, forthcoming-b Eliassmith 2007; Hohwy 2007.

Prediction-driven processing is processing that is driven by an imperative to reduce error in the brain's own (sub-personal) predictions concerning its current and future states. Importantly, it is the brain's attempt to predict its own *current* states that does much of the heavy lifting here. To see why this is crucial, it helps to think of the brain as a kind of black box that needs to learn about the world while having direct access (in the *processing* sense of direct access: I'll briefly return to more metaphysical notions of directness and access in sections 4 and 5) only to the states and changes of its own internal registers[4]. This image reminds us that the world, to put it bluntly, does not get to be inside the brain. Under such conditions, one good strategy is to deploy a learning routine that seeks to predict, on the basis of the current state of some higher neural population, the current state of a lower population. Each layer of neural processing is thus trying to predict the current response at the layer below (except for the bottom layer, such as the retina, which transduces some energetic signal). Under these conditions the brain can self-supervise its own learning, as the lower levels do indeed come to occupy sequences of states that the higher levels can (in the processing sense) access. That means that familiar forms of gradient descent learning can be applied so as gradually to move each higher level population closer and closer to settings (such as the values of synaptic weights) that actually enable successful prediction of the evolving states of the population below. By applying this kind of routine within a stacked hierarchy of neural populations, the brain comes to learn an inter-animated set of models (technically, it is then a single generative model[5]) that best capture the regularities that determine the shape of those evolving sensory signals. In the case of exteroception, these regularities involve the changing states of the external world. (For a simple early example of the successful use of such a routine, see Rao and Ballard 1999. For a more recent example, see Hohwy et al 2008).

[4] For a wonderful neuroscientific exploration of this issue, see Rieke at al 1999.

[5] Thanks to Karl Friston for pointing this out.

Embodiment and action fit very naturally within such a framework. Embodiment matters because embodied agents are active agents, and active agents can systematically alter their own sensory input streams in ways that can drive faster and more successful prediction-based learning. In particular, embodied agents can engage in what Pfeifer et al (2007) nicely dub the 'self-structuring of information flows' – for extensive discussion see Clark 2008. An additional reason that such active 'information self-structuring' may be computationally useful is because self-produced action applies a known (where that really just means repeatable-on-demand) transform to the incoming sensory data. Thus I can look at several scenes and in each case turn my head in just such and such a way. The fact that I am then applying the same kind of 'known transform' turns out to greatly facilitate the extraction of object-specifying information from the time-varying stream of sensory data (see Hinton et al 2011).

In addition, embodied agents have two deeply computationally related ways to reduce the errors in their own neural predictions. One way (perception) is to find and apply the right hypothesis to explain the current sensory signal. But the other way (action) is to issue a motor command that alters the sensory input bringing it in line with the prediction so as to reduce the error (see Friston 2009, 2010; Friston et al 2011). Both methods work together seamlessly in the embodied agent's constant quest to reduce prediction error within a hierarchy of sensorimotor processing regions. Perception and action thus emerge as two sides of a single computational coin.

Prediction-driven processing schemes, operating within hierarchical regimes of the kind just described, learn probabilistic generative models (within each higher neural population) of the activity patterns likely to be displayed in the neural population below. What is crucial here – what makes such models *generative* – is that once learnt they must be capable of predicting the patterns characteristic of (the responses to sensory input at) the level below. This means being able to generate, from the top-down (via a kind of step-wise downwards cascade) the activity patterns characteristic of each lower layer. The upshot is that such systems *simply as part and parcel of their ability to perceive* must develop an ability to self-generate perception-like states. In learning, such systems typically use perception (attempts at bottom-up detection) and generation (attempts at top-down prediction) in mutually bootstrapping ways. The upshot is that problems that at first sight look like problems of passive perception (recognizing and categorizing shapes, to take an example treated in Hinton 2007a) thus turn out to require developing something potentially much more active (an ability to generate 'images' of shapes). This deep mutuality between perception and (a kind of) imagination is striking, and I shall return to it shortly[6].

To sum up, converging paradigms[7] in neuroscience and machine learning suggest that perception is inextricably tied up not just with action but also with something functionally akin to imagination. They suggest, moreover, that prediction-driven processing, operating in a rich hierarchical setting within the brain, plays a major role in the development and deployment of perceptual understanding, and that self-generated action enables and enhances such learning in a variety of important ways.

4. The Dialogue, Revisited.

I claim that the picture just sketched accounts for everything that the enactivist points to as positive evidence for their more 'interactive' conception of experience itself, that it does so in a way that captures some key enactivist insights (concerning the important roles of prediction and self-generated action), and (further) that it fluently accommodates a number of personal-level phenomena that the pure

[6] I pursue this matter in more depth in Clark forthcoming-a.

[7] It is worth stressing that important differences separate many of the treatments reviewed above. What unites them is the use of prediction-based learning to acquire generative models within hierarchical processing regimes.

enactivist struggles valiantly to accommodate. I'll try to make good on each of these claims in turn, before ending with a few words about some remaining issues: ones that might reasonably be classed as more properly metaphysical in nature.

Let's start with the alternative account of the positive evidence for the enactivist conception of experience itself. That evidence, as we saw, comprised two main elements: some considerations concerning 'presence-in-absence', and various experimental results revealing variable neural correlates for similar experiences.

The alternative account of the first of these, presence-in-absence, is mostly straightforward. To perceive the world, according to the kinds of approach sketched in section 3 above, is (as the system relaxes into a stable state) to meet incoming driving sensory signals with matching top-down expectations generated using a hierarchy of increasingly abstract generative models. Basically, these will be models capturing regularities across larger and larger temporal and spatial scales. Inevitably, the higher levels here will come to encode information concerning the kinds of properties and feature that Noë highlights. Just as the higher levels in a shape recognition network respond preferentially to invariant shape properties (such as squareness or circularity) so we should expect to find higher level networks that model driving sensory inputs (as filtered via all the intervening levels of prediction) in terms of tomatoes, cats, and so forth. The overall processing hierarchy, confronted with a scene involving cat or tomato, will relax into a stable state in which these higher-level patterns are recognized to be present. The fact that only one side of the tomato is currently facing us, or that the cat (to use a recurring example – see e.g. Noë 2004) is partially occluded behind a picket fence, are precisely the kinds of additional detail that the higher level models must learn to ignore.

Noë writes that:

> The challenge of the theory of perception is to appreciate how perception can be, in this way, an encounter with how things are, when the nature of things necessarily exceeds what can be taken in at a glance" (Noë 2008 p. 691, and quoted in Ward 2012 p.??)

But this poses no problem whatsoever for accounts that must meet incoming sensory signals by relaxing into a cascade of top-down predictions involving, at the higher levels, models of distal causes such as chairs, tables, tomatoes, and cats. On the contrary, insofar as there is *any* kind of puzzle here for the alternative (prediction-and-generative-model based) account it concerns not 'presence-in-absence' but (paradoxically) 'absence-in-presence'! The puzzle, that is, is why we do not then *only* experience the cat/tomato as whole. This matters since, as Noë rightly insists, such perspectival information is important for many purposes, is fully consistent with (indeed, is handily diagnostic of) the correct higher-level story, and is clearly preserved (as experience attests) even once it has been 'used' to select the right higher-level hypothesis. Fortunately, there are various ways in which this may be achieved[8], perhaps the simplest of which is to assume that experience is conditioned upon the best linked set of hypotheses spanning multiple spatial and temporal scales (given current context and accommodating the driving sensory signal).

The situation with regard to variable neural correlates is, if anything, even more clear-cut. Given that the alternative (prediction-and-generative-model based) account posits a kind of canonical cortical microcircuit involved in all forms of sensorimotor learning, the potential recruitment of 'auditory' areas for visual processing (in the re-wired ferrets) and the potential for tactile visual sensory

[8] One recently suggested option involves the use of a more complicated architecture in which small processing ensembles encode "both the probability that the entity is present…and a set of "instantiation parameters" that may include the precise pose, lighting, and deformation of the visual entity relative to an implicitly defined canonical version of that entity" (Hinton et al 2011 p.2)

substitution emerge as natural consequences. The fact that such re-purposing occurs consequent upon some period of interactive engagement reflects only the mundane fact that prediction-driven learning requires the use of gradient-descent methods. Using such methods, predictions made at all levels of a processing hierarchy are gradually altered so as better and better to account for (to predict) the evolving sensory input stream across multiple spatial and temporal scales. Action, as we saw, plays a very special role in this, since self-generated actions provide a neat way of regularly manipulating the data stream so as to resolve ambiguities, speed up learning, and acquire the kinds of knowledge most needed for our species-typical practical engagements with the world. Many of the key enactivist insights concerning the role of action in perception are thus preserved within the alternative account.

The alternative account on offer does not, however, simply accommodate the same range of data and intuitions as the enactivist account. I think it does appreciably better. Thus recall that a major part of the argumentative apparatus that Ward brings to bear on the exchange with the enactivist concerns the need to achieve some form of 'fit' between personal-level experience and the suite of enabling sub-personal facts and mechanisms. The enactivist, as presented by Ward, uses the positive considerations just rehearsed to motivate a conception of experience as "essentially episodes of interaction between a subject and parts of the world' (Ward 2012 p. ??). But this conception has a well-known cost. It forces the theorist to take complex steps (this can be something of an understatement) to accommodate familiar and apparently highly relevant experiential phenomena, such as dreams and mental imagery. Time precludes a proper examination of the many ingenious ways that Noë and others hope to accommodate such phenomena while yet insisting that what experience actually *is* is ongoing dynamic interaction with the world (see eg Noë 2006 p 430-431, and discussion in Block 2005). I note only that a common move is to concede that dream experiences, unlike normal daily percepts, do indeed "depend on neural states alone" (Noë 2006 p. 431) but then to insist that they are quite unlike normal perceptual experiences that are instead "anchored by our dynamic coupling to the world" (Noë, op cit). Here, I think our alternative account can do significantly better, clarifying both the commonalities and the differences, in a way that makes more satisfying contact with ordinary experience.

Let's take the commonalities first. It is suggestive that early explorations of the use of generative models for perception deployed learning algorithms with names such as the 'wake-sleep algorithm' (Hinton et al 1995) and spoke easily and repeatedly of neural networks generating patterns for themselves 'in fantasy'. As we saw, where there is a generative model, at level N, of some pattern of response as it might occur at level N minus 1, there is the capability to actively promote (from the top down) that very pattern of activity at the lower level. Systems that know how to perceive an object as a cat are thus systems that, ipso facto, are able to use a top-down cascade to bring about the kinds of activity pattern that would be characteristic of the presence of a cat. Such systems thus display what (Clark (forthcoming-a)) I call a 'duality of perception and imagination'[9].

This seems to me to be of profound interest. Perceivers like us, if this is correct, are inevitably potential dreamers and imaginers too. Moreover, they are beings who, in dreaming and imagining, are deploying many of the very same strategies and resources used in ordinary perception. I think this (partial) mimesis between perception, dreaming, and imagining actually sits much better with our daily personal-level experience than do alternative accounts that treat them as radically divergent states.

[9] Note that this does not imply that subjects can bring this about at will, nor that their dreamings and imaginings will, in the typical case, reproduce ordinary perceptual states in all their detail and stability. For the former caveat, see Clark (forthcoming-a). For the latter, see the comments in section 4 following.

It is important, however, to do justice not just to the (at least apparently[10]) experienced commonalities but to the experienced differences too. Noë's response to the dreaming example, as mentioned above, is to suggest that dreams constitute a radically altered form of experience. This seems to me already to represent a major concession, since it allows that at least *some* genuine experiences are not well-characterized using the enactivist conception[11]. But it is, in any case, a move that the kind of prediction-and-generative-model-based internalism sketched earlier need not (and ought not) to reject. For nothing in that view requires that the system, when simply cycling, in the sleep or imagining state, in the absence of ongoing driving external inputs, will typically support the very same kinds of stability and richness of experienced detail that daily sensory engagements offer. Indeed, there are several reasons to doubt this.

In the absence of the driving sensory signal, no stable ongoing information about low-level perceptual details is there to constrain the system. As a result, there is no obvious pressure to *maintain* a stable hypothesis at the lower levels of processing: there is simply whatever relatively amorphous pressure the current higher-level story exerts. In waking life, we (or rather, our brains) can indeed expect - just as the enactivist stresses - the persisting external scene to provide vital stabilizing pressure. Moreover, altered chemical states of the brain accompany sleeping and there exists a surprising amount of overlap between the electrophysiological and neurochemical changes characteristic of the sleep-state and changes characteristic of schizophrenia and drug-induced hallucinations. Explanations of these

[10] The caveat recognized the logical space for a variety of disjunctivist proposals hereabouts, a consideration of which is well beyond the scope of the present treatment. For an excellent introduction, see the essays in Haddock and Macpherson 2008.

[11] An alternative would have been to insist that dream content inherits some kind of constitutive world-involvingness from ordinary perceptual content. For some useful discussion of the range of possibilities hereabouts, see Block 2005.

latter conditions[12] using the apparatus of generative models and hierarchical predictive processing (see e.g. Fletcher and Frith 2009 and Corlett et al 2009) are thus suggestive with regard to the peculiarities of sleep and dreaming. Thus Fletcher and Frith speculate that:

> "Perhaps the dream state arises from disruptions in hierarchical…processing such that sensory firing is not constrained by top-down prior information and inferences are accepted without question owing to an attenuation of the prediction-error signal from lower to higher levels" (Fletcher and Frith 2009 p. 52)

Much more work needs to be done to make this kind of suggestion precise. The point for current purposes is simply that the kinds of internalist model highlighted above[13] have the pleasing property of presenting perception, dreaming, and imagination within a single unifying framework while leaving plenty room for the kinds of typical difference (with respect to detail, richness, stability, and overall coherence) highlighted by Noë and the enactivists.

---

[12] Under conditions where perceptual detail matters, and where (the brain's) confidence in the driving sensory signal is high, the weighting on the prediction error units closer to the sensory peripheries is increased so as to ensure those error signals get to play a larger role in determining the current set of linked (multi-scale) hypotheses (see the discussion of 'precision' in Friston 2009 and in Feldman and Friston 2010). The flexible weighting of prediction error is a major feature of these models, and plays a central role in recent accounts, using this framework, of delusions and hallucinations in schizophrenia. The central hypothesis (Fletcher and Frith 2009) is that in such cases prediction error signals are falsely generated, wrongly weighted, and (thus) accorded undue salience (perhaps due to abnormal dopaminergic functioning).

[13] In Clark 2009 I also presented some important empirical considerations concerning the possible role of very fast temporal synchronies in the construction of coherent percepts. Such synchronies, it has recently been suggested, may play a crucial role in the flexible control of the influence (weighting) of prediction error. For discussion, see Feldman and Friston 2010.

5. The Porous Perceiver

Ward's treatment is sprinkled, as mentioned earlier, with another kind of consideration. Here, the enactivist is paired with philosophers such as McDowell (1986) and Putnam (1999) who are said to be opposing "the Cartesian conception of our minds as seats of pure subjectivity, blocked off from the environment by our sensory transducers" (Ward, 2012, p ??). The enactivist's conception, by depicting experience as essentially interactive and world-involving, is said to avoid this threat. It does so by building the experienced world right into the account of experience itself: an account that is now first and foremost an account of inescapably world-involving relations and interactions.

To adopt such a story, on the grounds that it is simply required if we are to ensure proper contact with the world (to ensure that perception can be genuinely world-revealing) is to pursue what, in the original article, I referred to as a 'more metaphysical' argument that might be thought to support something like ECM. In Clark 2009 I deliberately bracketed those arguments first, because they seemed to me to raise a host of quite different issues, and second, because the enactivist presented herself as offering us a set of novel considerations (concerning the functional role of action in perception) rather than as appealing to these more familiar forms of argument. My concern was thus only to engage with (and reject) the apparently novel considerations as offering good grounds for ECM or indeed (as we have just seen) for the enactivist conception of experience itself. Nonetheless, it may help to say a word about why the more metaphysical route doesn't actually strike me as in any way undermining the kind of empirical story I have been sketching.

The question to ask is, must an internalist account of (what we may call) the implementing machinery of conscious experience reject the idea that perception, when all goes well, is in some sense directly world-revealing? It is by no means obvious that it must. The kinds of mechanism I have sketched provide a means by which an agent may come to perceive the world. They do not, for example, provide a means by which an agent comes to perceive her own internal representations of the world. For what is perceived are not representations but the world. Representations, as they figure in cognitive scientific accounts of the kind on offer, are causally implicated in this process. And understanding this fact helps us to make empirical sense of the very possibility of dreaming, hallucination, and the like. Still, the super-tight empirical link (briefly sketched above) between the mechanisms of perception and the mechanisms of dreaming, imagination, and hallucination suggests that we should consider these latter cases as co-arising within the general setting of mechanisms for veridical perception and effective action. Considered in this way, the mechanistic story on offer simply shows *how* beings like us are able to achieve genuine access to the causal structure of our environment[14]. They show, as my colleague Matt Nudds once put it to me, how it is that we can be 'open to the world'. To the extent that this is correct, mere openness to the world provides no metaphysical trump card for the enactivist. Equipped with brains like ours we become porous to the world. Its structure and statistical regularities flow through us in as real a way as do food and water through the digestive tract.

6. Conclusions: Let The World In, and Dream

Ward (2012) invites us to reconsider what Clark (2009) presented as the enactivist arguments for the extended conscious mind (ECM), depicting them as first and foremost an invitation to embrace a world-involving conception of experience itself, and only in that way as (then more or less trivially) supporting an extended mechanistic basis for experience. The trouble with this is threefold.

[14] For a detailed argument consonant with this claim, see Millar 2007. Ward (2012, footnote 7) suggests that the enactivist, by contrast, cannot embrace Millar's proposal.

First, the positive considerations that Ward highlights are (or so I have argued) equally well accommodated by alternative (prediction-and-generative-model based) accounts. These accounts look, however, to be internalist about the mechanisms of perceptual experience. Second, these models have the added benefit of bringing perception, dreaming, and imagination into a common framework. This, I argued, is an advantage as it enhances the 'fit' (between personal-level experience and our best sub-personal stories) to which Ward centrally appeals. Finally - though this is a matter in need of much further discussion - the kinds of models I describe seem to me to deliver a means by which we might nonetheless properly be said to be perceptually open to the world.

Department of Philosophy, ANDY CLARK
School of Philosophy, Psychology, and Language Sciences,
University of Edinburgh,
EH8 9AD, UK

References

Bach y Rita, P and Kercel, S.W. 2003: 'Sensory Substitution and the Human-Machine Interface' *Trends in Cognitive Sciences* 7:12: 541-546

Ballard, D.H. 1991: 'Animate Vision' *Artificial Intelligence* 48, 57-86.

Block, N. 2005: 'Review of Alva Noë, *Action in Perception*', *Journal of Philosophy,* 5, pp.259-272

Bubic A, von Cramon DY and Schubotz RI. 2010: 'Prediction, cognition and the brain'. *Front. Hum. Neurosci.* 4:25: 1-15

Churchland, P.S, Ramachandran, V., & Sejnowski, T. 1994 : 'A Critique of Pure Vision'. In C. Koch & J. Davis (Eds.), *Large-Scale Neuronal Theories of the Brain* Cambridge, MA: MIT Press.23-61

Clark, A. 1997: *Being There: Putting Brain, Body, and World Together Again*, Cambridge, MA, MIT Press.

--2006: '"Cognitive Complexity and the Sensorimotor Frontier" *Proceedings Of The Aristotelian Society*, Supp. Vol LXXX 43-65

- - 2008: *Supersizing the Mind: Embodiment, Action and Cognitive Extension*, NY, Oxford University Press

— 2009: 'Spreading the Joy? Why the Machinery of Consciousness is (Probably) Still in the Head', *Mind*, 118, pp.963-993

-- forthcoming-a: 'Perceiving as Predicting' in M.Mohan, S. Biggs, and D. Stokes (eds) *Perception And Its Modalities* NY, Oxford University Press

-- forthcoming-b: 'Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science' *Behavioral and Brain Sciences*

Clark, A. and D. Chalmers 1998: 'The Extended Mind', *Analysis*, 58, pp. 7-19

Clark, A and Grush, R. 1999: 'Towards a Cognitive. Robotics' *Adaptive Behavior* 7 : 1: 5-16.

Corlett P., Frith C., Fletcher P. 2009: 'From drugs to deprivation: a Bayesian framework for understanding models of psychosis.' *Psychopharmacology* 206:4: p.515-30

Dayan, P., Hinton, G. E., & Neal, R. M. 1995: 'The Helmholtz machine'. *Neural Computation*, 7, 889–904.

Eliasmith, C. 2007: 'How to build a brain: From function to implementation'. *Synthese*. 153(3): 373-388

Feldman H., and Friston K. 2010: 'Attention, uncertainty, and free-energy'. *Frontiers in Human Neuroscience* 2:4:215.

Fletcher, P and Frith, C. 2009: 'Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia' *Nature Reviews: Neuroscience* 10: 48-58

Friston K. 2005: A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci*.29;360(1456) :815-36.

Friston K. 2009 : 'The free-energy principle: a rough guide to the brain?' *Trends in Cognitive Science* 13: p.293–301

Friston K. 2010: 'The free-energy principle: a unified brain theory?' *Nature Reviews: Neurosci*ence 11:2 p.127-38.

Friston K, Mattout J, and Kilner J. 2011: 'Action understanding and active inference'. *Biological Cybernetics*. 2011 104:137–160

Gregory, R. L. 1980: 'Perceptions as hypotheses'. *Phil. Trans. Royal Soc. Lond., Series B, Biological Sciences* 290 (1038): 181-197.

Grush, R. 1997: 'The Architecture of Representation'. *Philosophical Psychology* 10(1)5-23.

Grush, R. 2004: "The emulation theory of representation: motor control, imagery, and perception". *Behavioral and Brain Sciences* 27:377-442

Haddock, A., and Macpherson, F. (eds.) 2008: *Disjunctivism: Perception, Action, and Knowledge* Oxford: Oxford University Press).

Huang, Y. and Rao, R. 2011: 'Predictive coding'. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2 doi

Hinton, G. 2007a: 'To recognize shapes, first learn to generate images'. P. Cisek, T. Drew and J. Kalaska (Eds.) *Computational Neuroscience: Theoretical Insights into Brain Function*. Elsevier, Holland

Hinton, G. 2007b: 'Learning Multiple Layers of Representation'. *Trends in Cognitive Sciences,* 11, 428-434.

Hinton, G., Dayan, P., Frey, B. J., and Neal, R. M. 1995: 'The wake-sleep algorithm for unsupervised neural networks'. *Science,* 268: p.1158-1161

Hinton, G, Dayan, P, Neal, R & Zemel, R. 1994: 'Using neural networks to learn intractable generative models'. Invited paper at the 1994 meeting of the *American Statistical Association,* Toronto, Canada.

Hinton, G. E., Krizhevsky, A. and Wang, S. 2011: 'Transforming Auto-encoders'. paper presented to *ICANN-11: International Conference on Artificial Neural Networks*, Helsinki, Finland

Hinton, G., Osindero, S. and Teh, Y. 2006: 'A fast learning algorithm for deep belief nets'. *Neural Computation* 18, pp 1527-1554

Hinton, G. and Zemel, R. 1994: 'Autoencoders, minimum description length and Helmholtz free energy'. In Cowan, J. D., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems* 6: 3-10. San Francisco, Morgan Kaufmann

Hohwy, J. 2007: 'Functional Integration and the mind' *Synthese* 159:3: 315-328

Hohwy, J., Roepstorff, A., Friston, K. 2008: 'Predictive coding explains binocular rivalry: an epistemological review'. *Cognition* 108 (3): 687-701.

McDowell, J. 1986: 'Singular Thought and the Extent of Inner Space' in Philip Pettit and John McDowell, (eds.) *Subject, Thought, and Context*. Oxford: Clarendon Press, pp.137–68. References are to the version in

McDowell 1998

Millar, A. 2007: 'What the Disjunctivist is Right About', *Philosophy and Phenomenological Research,* LXXIV, 1, pp.176-198

Neisser, U. 1967: *Cognitive Psychology.* New York, Appleton-Century-Crofts

Noë, A. 2004: *Action in Perception,* Cambridge MA, MIT Press

—2006: 'Experience Without the Head', in Gendler and Hawthorne (eds) 2006: *Perceptual Experience*, New York, Oxford University Press, pp.411-434

Noë, A. and J.K. O'Regan 2001: 'A Sensorimotor Theory of Vision and Visual Consciousness', *Behavioural and Brain Sciences,* 24, pp.939-1031

Pfeifer, R, Lungarella, M., Sporns, O., and Kuniyoshi, Y. 2007: 'On the information theoretic implications of embodiment- principles and methods'
Lecture Notes in Computer Science 4850, p. 76-86

Putnam, H. 1999: *The Threefold Cord: Mind, Body and World*, New York, Columbia University Press

Rao, R and Ballard, D. 1999: 'Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects' *Nature Neuroscience* 2, 1, 79

Rieke, F., Warland, D., van Steveninck, R. and Bialek, W. 1999: *Spikes: Exploring the Neural Code* Cambridge, MA, MIT Press

Sur, M., Angelucci, A., and Sharma, J. 1999: 'Rewiring cortex: the role of patterned activity in development and plasticity of neocortical circuits' *Journal of Neurobiology* 41: 1: 33-43

Ward, D. 2012: 'Enjoying the Spread: Conscious Externalism Reconsidered' *Mind* this issue.